

## **LINEALIDAD, MONODIMENSIONALIDAD E INVARIANZA. REQUISITOS DE LOS TESTS**

*Miguel Ángel Rosado Chauvet*

### **RESUMEN**

*Dos de las características que deben poseer los instrumentos de medición son la linealidad y la monodimensionalidad.*

*Cualquier instrumento que se utilice para medir conocimientos, valores afectivos o habilidades motoras debe describir con precisión la dimensión del constructo por evaluar con el objeto de delimitar el campo conceptual, afectivo o psicomotor de que se trate. Posteriormente, debe elaborarse una cantidad suficiente de ítems que cubran una cantidad representativa de los aspectos que se pretende medir, tomando como base una tabla de especificaciones en la que se incluya tanto el contenido como el nivel taxonómico que se espera medir. Finalmente, se elige una cantidad suficiente y representativa del total de ítems elaborados, después de realizar una aplicación piloto en la que se someten a un análisis previo a la aplicación definitiva.*

\* Departamento de Economía de la UAM-I.

*Las mediciones deben ser isomórficas entre habilidades de los sujetos y dificultades de los ítems, donde los sujetos con menores habilidades ocupen los valores menores del continuo y los sujetos con habilidades mayores se sitúen en los valores mayores.*

*En caso de tener más de un tipo de contenidos, los análisis de ítems deben realizarse por secciones temáticas en vez de hacerlo por el instrumento en su totalidad, para evitar la contaminación entre las secciones. Una persona puede tener habilidad suficiente en uno de los temas y una habilidad deficiente en otro y, si se incluyen ambos en un solo constructo, el resultado puede ser una habilidad intermedia.*

*Aun cuando la calificación final incluya a todos los sujetos, para realizar el análisis de ítems debe incluirse sólo a los sujetos que se encuentren entre los límites probabilísticos aceptados, de manera que la posesión o carencia del constructo por medir no influya en el análisis.*

*No se deben aplicar los mismos criterios para evaluar sujetos que para evaluar ítems. En los ambientes educativos, los límites para aceptación de los sujetos están dados por el total de aciertos al instrumento, como límite superior, y por el cupo de la institución, como límite inferior. En las empresas se requiere que los sujetos cubran un mínimo indispensable (criterio) de habilidades y conocimientos para el puesto que se solicita.*

*Por otra parte, cuando se evalúan los ítems deben establecerse un mínimo y un máximo admisibles de dificultad, así como una discriminación mínima que permita saber si el ítem permite admitir a los sujetos que tengan la habilidad o los conocimientos suficientes para continuar sus estudios. No se trata de discriminar a los sujetos por características de género, religión, raza, nacionalidad u otras, sino por capacidades de obtención de los conocimientos y habilidades requeridos para desarrollar un trabajo con eficiencia. Hay que tener en cuenta que la sociedad espera que los egresados de una institución puedan solucionar los problemas para los cuales se les contrata.*

*El objetivo básico y directo de la teoría de respuesta al ítem (TRI) propone una solución planteado que a través del método logístico se obtienen mediciones invariantes con respecto a los tests, lo cual permite disponer de tests cuyas características son invariantes con respecto a los sujetos. Sin embargo, por tratarse de una curva acumulativa asíntota, hay que determinar los límites dentro de los cuales se pueda llegar a una variabilidad mínima.*

Palabras clave: isomorfismo, constructo, ítem, linealidad, invarianza, monodimensionalidad o unidimensionalidad, logístico, asíntota.

### **Linealidad en el modelo logístico**

Uno de los supuestos de los modelos logísticos se refiere a la linealidad de las proporciones de respuesta con respecto a una curva logística, la cual es una asíntota.

En el estudio que nos ocupa se relacionan los valores de la línea recta con la curva logística, con valores básicos del logaritmo natural de proporciones de acierto entre proporciones de error  $\ln(p/q)$ , con el propósito de verificar hasta qué punto podemos asumir dicha linealidad.

Un segundo aspecto estudiado corresponde a la diferencia probabilística que se presenta al utilizar ítems de diferentes opciones en vez de la forma tradicional de respuestas dicotómicas.

### **Método**

Este estudio corresponde a un análisis teórico sobre la relación entre la curva logística y la línea recta, por lo que no se utilizarán ni sujetos ni instrumentos. Los dos elementos teóricos bajo comparación son:

*Curva logística*

$$\ln = \frac{p}{1-p}$$

Linealidad, monodimensionalidad e invarianza. Requisitos de los test

### *Línea recta*

$$y = a + bx$$

### *Procedimiento*

Se utilizará la regresión entre las proporciones de acierto y error en valores de la curva logística y los valores de una recta con el método de incremento constante de los valores logísticos extremos.

El proceso se iniciará con los valores extremos de la logística, eliminándolos hasta lograr un máximo de 0.05 redondeando el excedente de error probable de medición tanto en la constante ( $a$ ) como en el valor de ponderación ( $x$ ) sobre la pendiente.

Otro valor importante se refiere a un coeficiente de determinación ( $r^2$ ) de +1.00 o próximo a este valor.

Como complemento del estudio se verificará qué tanto varían los valores en función del número de opciones de un ítem, variando las probabilidades de respuesta de acuerdo con los límites inferior (LI) y superior (LS), según la propuesta presentada para análisis de ítemes en documentos anteriores.<sup>1</sup>

## **Resultados**

### *Con respecto a la linealidad*

Para obtener el rango donde se encuentra la linealidad máxima, se procedió a obtener los lógitos básicos [ $\ln(p/q)$ ] para todas las proporciones entre 0.99 y 0.01, así como la línea entre los extremos mediante una proporción de decremento constante.

<sup>1</sup>Nos referimos inicialmente a “Cinco estudios de estadística para la evaluación”, publicado en los *Cuadernos del CENEVAL* en junio de 1996, así como a publicaciones internas posteriores.

La constante (K) es prácticamente de 0.000 y el error estándar de la variable “x” es menor que 0.05 (5%) para cualquiera de los casos presentados.

Se observa que un lógito básico de  $\pm 2.000$  ( $\pm 1.992$ ) permite un error estándar de la variable “y” (ESy) del orden de 0.09 (0.088), lo que llega a ser de 0.05 (0.054) hasta un lógito de  $\pm 1.658$  con una proporción de  $p = 0.84$  y  $q = 0.16$  y con 0.997 de coeficiente de determinación ( $r^2$ ).

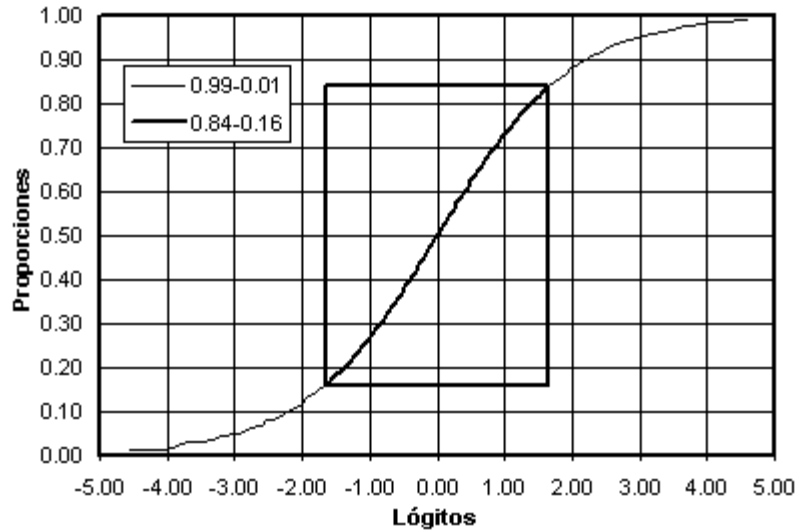
De lo anterior se concluye que para obtener una máxima linealidad admisible deben tomarse como valores proporcionales extremos  $p = 0.84$  y  $q = 0.16$ , con su lógito básico que corresponde a  $\pm 1.658$ , para mantener un 0.05 (5%) de error estándar de medición en los lógitos de acuerdo con la línea recta, con  $r^2 = 0.997$  de coeficiente de determinación.

Se notará que este rango es muy próximo a los valores “normales” de la curva normal ( $\pm 1.00$ ), desviaciones estándar (0.1587 $\leftrightarrow$ 0.8413), por lo que es válido para el 68.26% de los casos, y con valores de la curva logística que presentan una distancia aceleradamente mayor a medida que el rango entre “p” y “q” es mayor, hasta un 0.64 (64%) de error en la constante “K” y un 0.04 de error en el coeficiente “x” con  $p=0.99$  y  $q = 0.01$  de proporciones.

### Regresión entre las líneas logística y recta

<b>K</b>	0.000	0.000	0.000	0.000
<b>ESy</b>	0.641	0.114	0.088	0.054
<b>r<sup>2</sup></b>	0.944	0.992	0.994	0.997
<b>r</b>	0.972	0.996	0.997	0.999
<b>N</b>	99	81	77	69
<b>gl</b>	97	79	75	67
<b>x</b>	1.530	1.143	1.119	1.083
<b>ESx</b>	0.038	0.011	0.010	0.007
<b>p</b>	0.99	0.90	0.88	0.84
<b>q</b>	0.01	0.10	0.12	0.16
<b>Lógito</b>	4.595	2.197	1.992	1.658

**Puntos de máxima linealidad sobre la curva logística**



En el recuadro interior de la gráfica se encuentran los límites de proporción entre 0.84 y 0.16, así como los valores entre -1.658 [ln(0.16/0.84)] y +1.658 [ln(0.84/0.16)], que corresponden al rango de lógitos.

Estos mismos valores de 0.16 y 0.84 corresponden al cálculo de los límites inferior (LI) y superior (LS) del modelo SISEVAL<sup>®</sup> cuando las opciones de respuesta son 4 y se trabaja con proporciones de 100. Al utilizar 5 opciones las proporciones corresponden a 0.12 y 0.88, que coinciden con la aproximación a ±2.00 lógitos. En ambos casos los límites de SISEVAL se calculan sobre un total de 100 para tratarlos como proporciones, así como con un valor de z = 1.96 para a = 0.05 como error probable de medición.

*Índice de dificultad aplicado a los ítems*

$$I_{Dif} = \frac{A}{S}$$

$$LI = \frac{S}{S_k} - \frac{z\sqrt{S(k-1)}}{2S}$$

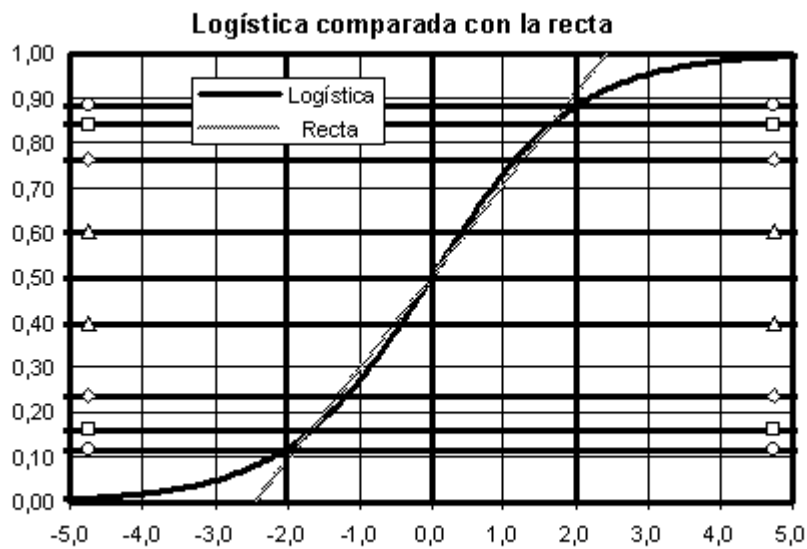
$$LS = \frac{S(k-1)}{S_k} + \frac{z\sqrt{S(k-1)}}{2S}$$

donde:

- A = Número de aciertos
- S = Número de sujetos
- K = Número de opciones del ítem
- z = Valor probabilístico normal (a = 0.05; z = 1.96)

Decisión: Si  $I_{Dif} < LI$  o  $I_{Dif} > LS$ , se rechaza el ítem.

Debe considerarse que para cantidades pequeñas de S, el rango de probabilidad entre los límites inferior y superior será mayor que para cantidades grandes, afectando los valores extremos en función de la variabilidad intrasujeto.



Rango entre los límites inferior y superior con varias cantidades de opciones:

- 5 opciones = círculos ○
- 4 opciones = cuadrados □
- 3 opciones = rombos ◇
- 2 opciones = triángulos △

### Aproximaciones a valores usuales

	<b>p</b>	<b>q</b>	<b>k</b>	<b>Lógito</b>
<b>a</b>	0.88	0.12	5	2.025
<b>b</b>	0.84	0.16	4	1.657
<b>c</b>	0.76	0.24	3	1.175
<b>d</b>	0.60	0.40	2	0.418

$p$  = Acierto;  $q$  = Error;  $k$  = Opciones; Lógito =  $\ln(p/q)$

- a) Valores aceptados por el método logístico y SISEVAL<sup>®</sup> como extremos (2.000)
- b) Valores de proporciones de normalidad de la curva normal ( $m \pm 1s$ )
- c) Valores cercanos al rango intercuartilar ( $Q_3 = 0.75$ ;  $Q_1 = 0.25$ )
- d) Valores de normalidad en puntuaciones estándar ( $m = 0.50$ ;  $s \pm 0.10$ )

### Conclusiones

Una curva logística totalmente discriminatoria dicotomiza en el eje de “ $x$ ” los lógitos negativos de los positivos; tanto unos como otros son una paralela a este eje con una vertical perpendicular al eje de “ $x$ ” y paralela al eje de “ $y$ ”. A medida que adquiere un desplazamiento hacia la derecha en la parte superior y un desplazamiento hacia la izquierda en la parte inferior, separándose de la perpendicular, adquiere también características de linealidad; sin embargo, también pierde discriminación hasta la situación extrema contraria, donde es paralela al eje de “ $x$ ” y perpendicular al eje de “ $y$ ”, coincidiendo con una proporción de acierto de 0.50 (50%) y error de 0.50 complementario.

Por lo anterior, tomar como valor extremo en forma irrestricta  $\pm 2.000$  atenta contra la discriminación de los *ítemes* en cualquier caso que no sea de 5 opciones de respuesta, y en el caso extremo de valores dicotómicos, cualquiera



de los lógitos básicos que exceda de  $\pm 0.418$  tenderá a violar la discriminación asociada con el punto de inflexión de la pendiente y, al ser un indicador no controlado por el modelo, puede favorecer puntajes de bajo rendimiento y viceversa.

Existen tres tendencias en cuanto a la solicitud de respuestas:

1. “Conteste a todas las preguntas aun cuando no las sepa.”
2. “Si no sabe la respuesta, conteste *no sé*.”
3. “Deje en blanco la respuesta si no está seguro.”

En el primer caso se propicia un factor de error adjudicable a las respuestas “por azar”, por lo que se debe tomar en cuenta una proporción no controlada por el modelo logístico básico (Rasch para un parámetro), el cual trabaja sobre valores dicotómicos sin tomar en cuenta las probabilidades de error asociadas al número de opciones de respuesta.

En el segundo caso se producen dos condiciones de error de medición:

- a) Si se califican las faltas de respuesta como “no respuesta”, se favorece a los sujetos al asignar valores de proporción de aciertos sobre respuestas, disponiendo de mayor tiempo de resolución en donde acierta, o
- b) si se califican las faltas de respuesta como “error”, nada garantiza que los sujetos no sepan lo que no contestaron, tomándose como “aciertos” respuestas a las que se dedicó mayor tiempo del disponible y como “errores” la falta de certeza sobre el desconocimiento de la solución.

Por último, en caso de responder “no sé” y tomar esto como respuesta para efectos del cálculo de proporciones de acierto, no sabemos si el desconocimiento es completo o discrimina entre varios aspectos de la pregunta pero no puede discriminar entre dos de las opciones que se le presentan como posible solución.

En cualquiera de los casos, se sugiere solicitar respuesta completa y controlar la varianza de error adjudicable al número de opciones, como posible control de la calidad de respuestas, aun cuando esto implique restar un puntaje real a los sujetos de mayor habilidad o capacidad, que responden en forma determinante y clara pero son “castigados” con un factor de “azar” que no es real para ellos. En todo caso, para efectos de admisión, certificación o promoción es más crítica la decisión entre los límites inferiores que entre los superiores.

### **Monodimensionalidad**

Cuando elaboramos un instrumento debemos definir los límites del *constructo* y realizar el análisis de ítems con la congruencia que requiera la especificidad de la evaluación.

A una institución puede resultarle importante un constructo global sobre determinado plan de estudios. En este caso podría efectuarse un análisis de ítem con respecto a la currícula general, por lo cual tiene sentido realizar el análisis entre ítem y test. Sin embargo, puede intentarse definir una preferencia específica sobre un grupo de materias con el propósito de impartir una orientación vocacional, en cuyo caso el análisis de ítems se debe realizar contra cada una de las materias de los planes de estudio.

Un instructor al que sólo le interese conocer el nivel de conocimientos de los participantes con fines de promoción, deberá realizar un análisis de ítems global sobre la temática general de su materia; en este caso es suficiente el análisis ítem-test con un test que integre un muestreo adecuado, suficiente y representativo de su programa de estudios. No obstante, si su objetivo consiste en determinar en forma específica los temas en que ha habido un mayor rendimiento, deberá utilizar un análisis ítem-tema. Esta característica sigue siendo vigente cuando utilizamos modelos más complejos, como el análisis factorial.

El estudio que presentamos señala la diferencia de los resultados cuando se trabaja con un análisis general y cuando se hace con uno específico, incluyendo un análisis adicional sobre los efectos de la muestra de datos utilizada, con muestras aleatorias o depuradas.

## **Método**

Trabajar con muestras seleccionadas al azar no garantiza por fuerza que todos los sujetos seleccionados deban ser tomados en cuenta al realizar un análisis de ítemes; se deben juzgar además las formas de respuesta que pueden sesgar los resultados. Con este propósito, se sugiere el análisis de la calidad de la muestra mediante *suficiencia de respuesta (SR)*, *pertinencia con respecto al constructo (PT)* y *pertinencia con respecto al contenido (PC)*.

En otro sentido, tal vez no sea adecuado tomar como parámetro el total del instrumento para realizar un análisis ítem-test si este último contiene temáticas diversas que no guardan relación entre sí, ya que se obtendrá una mayor frecuencia de respuesta de acuerdo con las preferencias y aprendizajes del sustentante. Por lo anterior, la sugerencia en este caso consiste en realizar análisis ítem-tema en función de la presentación de resultados por materia específica.

## *Sujetos*

Se tomó una muestra aleatoria de 90 casos para cada una de las 10 versiones del EXANI-I (Examen Nacional de Ingreso al Nivel Medio Básico), aplicadas en 1994 en el ámbito nacional, incluyendo todas las entidades federativas y las instituciones donde se aplicaron.

## *Instrumento*

El EXANI-I aplicado en 1994 contenía 120 ítemes de cinco opciones, divididos en 10 temas de 12 *ítemes* cada uno. Los “ítemes” se incluyeron cuidando la temática, así como las proximidades de los valores de los índices de dificultad y discriminación, por lo que se asume una equivalencia relativa aceptable.

## **Procedimientos**

### *Procedimiento “A” (tradicional)*

- Se incluye el índice de suficiencia de respuestas dividiendo el número de respuestas entre el número de preguntas del test, aun cuando esto no se utiliza en el método tradicional, con el objeto de verificar el impacto del procedimiento propuesto.
- Se calcula el índice de dificultad dividiendo el número de aciertos a un ítem entre el número de sujetos (se ignoran los límites críticos).
- Se calcula el índice de discriminación correlacionando las respuestas a un ítem con las respuestas al test y excluyendo los ítems que no lleguen al límite crítico inferior.

### *Procedimiento “B” (propuesto)*

Para los sujetos (depuración de la muestra)

- Se calcula el índice de suficiencia de respuesta de los sujetos dividiendo el número de respuestas entre el número de preguntas del test, y excluyendo a los sujetos que no lleguen al límite crítico inferior.
- Se calcula el índice de pertinencia con respecto al contenido dividiendo el número de aciertos entre el número de ítems del “test”, excluyendo los ítems que tengan una frecuencia de respuesta menor que el límite crítico inferior o mayor que el límite crítico superior.
- Se calcula el índice de pertinencia con respecto al constructo correlacionando las respuestas de un sujeto contra las respuestas del grupo en el test, y eliminando a los sujetos que no alcancen el límite crítico inferior.

Para los ítemes

- Se calcula el índice de suficiencia de respuesta al *ítem*, dividiendo el número de respuestas a un ítem entre el número de sujetos aceptados y eliminando a aquellos ítemes que no alcancen el límite crítico inferior.
- Se calcula el índice de dificultad, dividiendo el número de aciertos a un ítem entre el número de sujetos aceptados y excluyendo los ítemes que tengan una frecuencia de respuesta menor que el límite crítico inferior o mayor que el límite crítico superior.
- Se calcula el índice de discriminación, correlacionando las respuestas a un ítem con las respuestas al tema de los sujetos aceptados y eliminando los ítemes que no lleguen al límite crítico inferior.

Se verificó la *suficiencia de respuesta* en todos los sujetos, tomando como límite crítico (LC) el valor obtenido en la fórmula:

$$LC = \frac{N(k-1) - z\sqrt{N(k-1)}}{Nk} - \frac{1}{2N}$$

donde:

- N = Número de ítemes (120)  
 k = Número de opciones (5)  
 z = Valor probabilístico (1.96)

Decisión: Si la proporción de respuestas a los ítemes es menor que el límite crítico, se rechaza el caso.

Posteriormente se verificó la *idoneidad* de la muestra con respecto a los límites críticos mínimo (LI) y máximo (LS) de acierto, de acuerdo con las fórmulas:

$$LI = \frac{N - z\sqrt{N(k-1)}}{Nk} - \frac{1}{2N} \quad LS = \frac{N(k-1) + z\sqrt{N(k-1)}}{Nk} - \frac{1}{2N}$$

donde:

- N = Número de ítems (120)
- k = Número de opciones (5)
- z = Valor probabilístico (1.96)

Decisión: Si la proporción de aciertos a los ítems es menor que el límite crítico inferior o mayor que el límite crítico superior, se rechaza el caso.

Finalmente se verificó la *idoneidad* de la muestra con respecto a la congruencia con el grupo, utilizando la fórmula de correlación producto-momento de Pearson, tomando como límite crítico:

$$LC = \frac{z}{\sqrt{N}}$$

donde:

- z = Valor probabilístico (1.96)
- N = Número de ítems (120)

Decisión: Si el valor de correlación es menor que el límite crítico, se rechaza el caso.

Mediante este procedimiento se separaron los sujetos en dos grupos (“Aceptados” y “Rechazados”) para cada una de las 10 versiones del instrumento.

El índice de suficiencia de respuesta, tanto para los ítems como para los sujetos, permite que los análisis se realicen sobre datos más reales, eliminando la duda acerca de si la medición es sobre el objeto o sobre su desconocimiento.

Al depurar la muestra se garantiza que los datos utilizados para el análisis de ítems se basan en parámetros correctos, lo que da una mayor congruencia al estudio ya que se elimina la varianza adjudicable a los sujetos.

El establecimiento de límites críticos permite emitir juicios sobre la calidad de la muestra y de los ítemes en función de valores probabilísticos ampliamente admitidos en los métodos estadísticos.

Al realizar contrastes del ítem con respecto al tema al que corresponde, eliminamos la varianza originada por un parámetro sesgado. Sirve de aclaración un ejemplo extremo: si un sujeto es excelente en Español y pésimo en Matemáticas, en un examen de conocimientos que incluya ambos temas resultaría un sujeto mediocre si el parámetro para la correlación en el cálculo del índice de discriminación fuera el total de conocimientos. Pero si la evaluación del ítem se realiza por tema, el sujeto correspondería a su propia realidad como parámetro, resultando ser excelente o pésimo en función del tema que sirva de contraste al ítem.

## **Resultados**

El primer análisis que reportamos (ver Anexo A) consiste en dividir la muestra en “aceptados” y “rechazados”. Al primer grupo corresponden los sujetos que no cubren los límites de decisión de uno o más de los métodos mencionados, mientras que el segundo grupo se integra con los sujetos que cubren los tres criterios. Del total de 900 casos, 179 fueron rechazados (19.9%), con una oscilación de entre 6 (6.7%) y 32 (35.6%) casos rechazados entre las versiones; los casos aceptados en el total son 721 de 900 (80.1%), oscilando entre 84 (93.3%) y 58 (64.4%), los cuales son complementarios de los anteriores en las versiones 5 y 9 respectivamente.

Se hace notar que el promedio de variabilidad del grupo “rechazado” es de 0.925 y el de “aceptados” es de 0.310, lo cual indica una disminución en la varianza adjudicada a la muestra que, de no controlarse, tendrá un efecto sobre el análisis de ítemes que se base en ella, sesgando los resultados.

En el segundo análisis (ver Anexo B), partiendo de 120 ítemes del grupo no depurado y sin eliminar ningún ítem, el promedio de respuestas es de 0.986, con 0.389 de índice de dificultad y 0.244 de índice de discriminación. Cuando se procede a depurar al grupo, aun cuando no se elimine ningún ítem el promedio de

respuestas es de 0.995, con 0.408 de índice de dificultad y 0.374 de índice de discriminación.

Al realizar el análisis de ítems se observa que con el grupo no depurado se admiten 76 de los 120 ítems (63.3%), con un promedio de respuestas de 0.986 y con 0.409 de índice de dificultad y 0.326 de índice de discriminación. Trabajando con la muestra depurada se admiten 104 de los 120 ítems (86.7%), con un promedio de respuestas de 0.994 y con 0.417 de índice de dificultad y 0.409 de índice de discriminación.

## **Conclusiones**

*En las 10 versiones*

Método “A”

- Se aceptan 76 (63.3%) de los 120 ítems al contrastarlos contra el test.
- Se mantiene la misma cantidad de 90 (100.0%) sujetos.
- La suficiencia de respuesta (SE) se mantiene igual (0.986).
- El índice de dificultad se incrementa en 0.389 a 0.409 al depurar los ítems.
- El índice de discriminación se incrementa de 0.244 a 0.326 al depurar los ítems.

Método “B”

- Se aceptan 104 (86.7%) de los 120 ítems al contrastarlos contra el tema.
- La cantidad de sujetos disminuye de 90 a 72 (80.0%) al depurarlos.
- La suficiencia de respuesta (SE) disminuye de 0.995 a 0.994.



- El índice de dificultad se incrementa de 0.408 a 0.417 al depurar los ítemes.
- El índice de discriminación se incrementa de 0.374 a 0.409 al depurar los ítemes.

**Diferencia entre métodos**

	<b>I</b>	<b>S</b>	<b>SE</b>	<b>Dif</b>	<b>Dis</b>
<b>AT</b>	120	90	0.986	0.389	0.244
<b>AD</b>	76	90	0.986	0.409	0.326
<b>BT</b>	120	90	0.995	0.408	0.374
<b>BD</b>	104	72	0.994	0.417	0.409
<b>BD-AT</b>	-16	-18	0.008	0.028	0.166

### **Variabilidad en el método logístico**

En la teoría clásica las mediciones en las variables son dependientes del instrumento que se aplique para medirlas. Esta variable es definida operativamente por el instrumento con el cual se mida. Además, los ítemes y por ende el test donde se integran, dependerán de los sujetos a los que se aplican.

El objetivo básico y directo de la teoría de respuesta al ítem (TRI) propone una solución para este problema, permitiendo:

1. Obtener mediciones que sean invariantes con respecto a los tests.
2. Disponer de tests cuyas características sean invariantes con respecto a los sujetos.

Además, la TRI proporciona un conjunto de técnicas para la evaluación psicológica, como las *funciones de información* de los ítemes y del test, *los errores típicos de medida* distintos para cada nivel de la variable y el establecimiento de *bancos* de ítemes con parámetros estrictamente definidos, lo cual permite utilizar tests específicos para el nivel del sujeto, así como evaluaciones más completas y exactas en función de los rasgos del sujeto evaluado.

## **Problema**

Si los ítemes miden temas idénticos, las dificultades son equivalentes, la habilidad de los sujetos es similar –siempre y cuando se hayan seleccionado en forma aleatoria del total de aplicaciones– y las condiciones de aplicación corresponden a una situación similar de admisión, entonces deben presentarse los supuestos del método logístico con respecto a la invarianza de resultados con independencia del instrumento y de los sujetos que lo responden.

Sin embargo, ¿existirá la misma invarianza al hacer variar las condiciones mencionadas en los sujetos o en los tests aplicados?

Si se aceptan como reales los postulados de la TRI en condiciones normales, ¿qué ocurrirá en los resultados de los sujetos evaluados al someterse a condiciones extremas?

Utilizando tests integrados con la misma temática y de acuerdo con una sola tabla de especificaciones, aplicados bajo las mismas condiciones utilizando muestras aleatorias y tomando: a los mismos sujetos, con “tests” similares y dificultades similares; a los mismos sujetos, con “tests” similares y dificultades diferentes; a sujetos de diferente nivel de habilidad, con “tests” similares y dificultades similares; y a sujetos de diferente nivel de habilidad, con “tests” similares y dificultades diferentes.

## **Hipótesis**

Dado que la correlación implica el orden jerárquico de los valores, una correlación positiva perfecta indica una relación directa (isomórfica) de los lugares que ocupa cada uno de los casos. Mientras más bajo es el valor de la correlación, el orden de los datos varía más.

La invarianza de los resultados sugiere que, al margen de cuáles sean las dificultades de los ítemes que se elijan o la habilidad de los sujetos del grupo que responda al test, los resultados deben ser similares para la misma persona que responde.

Debe encontrarse muy próxima a +1.00 la correlación que reporten un test “fácil”, uno “medio” y uno “difícil” basados en una misma tabla de especificaciones (40 ítems con 10 temas diferentes), con respecto al total de “ítems” de los tres instrumentos (120 “ítems” con 10 temas diferentes) y con 10 versiones para cada test:

- En los resultados de los mismos sujetos con tests que requieren de habilidades “alta”, “media” y “baja”:

$H_1$  con independencia de las versiones que se utilicen.

$H_2$  dependiendo de las versiones aplicadas.

- En los resultados del mismo test:

$H_3$  con independencia de las versiones que se apliquen.

$H_4$  con independencia de la habilidad de los sujetos.

Dado que se aduce invarianza en cualquier condición, las correlaciones deben ser suficientemente altas para que los sujetos ocupen lugares similares con una distancia mínima en las condiciones utilizadas como propiciadoras de varianza, por lo que no basta que las correlaciones reportadas sean significativas a los niveles de probabilidad tabulares ( $N = 30$ ;  $r = 0.352$ ;  $z = 1.960$ ;  $\alpha = 0.05$ ), sino que establecemos como mínimo  $r = 0.800$  de correlación.

## **Método**

### *Sujetos*

Se obtuvo un muestreo aleatorio de cada una de las 10 versiones aplicadas en 1994 en el ámbito nacional, con 120 casos por versión, lo que hace un total de 1,200 sujetos.

### *Instrumentos*

Cada una de las versiones incluyó 10 temas diferentes, de acuerdo con los planes y programas de estudio de las características del nivel medio básico, con 12 ítems por cada tema.

### **Procedimiento**

#### *Para los sujetos*

Se obtuvieron muestras idóneas verificando que tuvieran suficientes repuestas para ser consideradas (*suficiencia de respuesta*), que los aciertos estuvieran dentro de un rango admisible (*pertinencia con respecto al contenido*) y que el sujeto contestara en forma congruente con el grupo al que pertenece (*pertinencia con respecto al tema*).

#### *Suficiencia de respuesta (SR)*

$$SR = \frac{R}{I}$$

$$LC = \frac{R(k-1) - z\sqrt{R(k-1)}}{Rk} - \frac{1}{2R}$$

donde:

- R = Número de respuestas
- I = Número de ítems
- k = Número de opciones del ítem
- z = Valor probabilístico normal ( $\alpha = 0.05$ ;  $z = 1.96$ )

Decisión: Si  $SR < LC$ , se rechaza al sujeto.

*Pertinencia con respecto al contenido (PC)*

$$PC = \frac{A}{R}$$

$$LI = \frac{I - z\sqrt{I(k-1)}}{Ik} - \frac{1}{2I} \quad LS = \frac{I(k-1) + z\sqrt{I(k-1)}}{Ik} - \frac{1}{2I}$$

donde:

- A = Número de aciertos
- I = Número de ítems
- k = Número de opciones del ítem
- z = Valor probabilístico normal ( $\alpha = 0.05$ ;  $z = 1.96$ )

Decisión: Si  $PC < LI$  o  $PC > LS$ , se rechaza al sujeto.

*Pertinencia con respecto al tema<sup>2</sup> (PT)*

$$PT = \frac{ISG - (S)(G)}{\sqrt{I S^2 + (S)^2 I + G^2 + (G)^2}}$$

$$r = \frac{z}{\sqrt{I - 1}}$$

donde:

- I = Número de *ítemes*
- S = Respuestas del sujeto
- G = Respuestas del grupo
- z = Valor probabilístico normal ( $\alpha = 0.05$ ;  $z = 1.96$ )

<sup>2</sup> Constructo.

Decisión: Si  $PT < r$ , se rechaza al sujeto.

Los criterios de selección de los sujetos son los siguientes:

1. Superior al SE (mientras más alto, mejor)
2. Entre LI y LS (mientras más cercano a 0.500, mejor)
3. Superior a  $r$  (mientras más alto, mejor)

Se eliminó a los sujetos que no correspondían a los límites críticos mencionados y se eligieron los 90 casos que respondían mejor a los criterios de selección.

*Para los ítems*

Se verificó que tuvieran suficientes repuestas para ser considerados (*suficiencia de respuesta*).

*Suficiencia de respuesta*

$$SR = \frac{R}{S}$$

$$LC = \frac{S(k-1) - z\sqrt{S(k-1)}}{Sk} - \frac{1}{2S}$$

donde:

- R = Número de respuestas
- S = Número de sujetos
- k = Número de opciones del ítem
- z = Valor probabilístico normal ( $\alpha = 0.05$ ;  $z = 1.96$ )

Decisión: Si  $SR < LC$ , se rechaza al sujeto.

Se obtuvo el índice de dificultad de cada uno de los ítemes.

$$IDif = \frac{A}{S}$$

$$LI = \frac{S - z\sqrt{S(k-1)}}{Sk} - \frac{1}{2S} \quad LS = \frac{S(k-1) - z\sqrt{S(k-1)}}{Sk} - \frac{1}{2S}$$

donde:

- A = Número de aciertos
- S = Número de sujetos
- k = Número de opciones del ítem
- z = Valor probabilístico normal ( $\alpha = 0.05$ ;  $z = 1.96$ )

Decisión: Si  $IDif < LI$  o  $IDif > LS$ , se rechaza el ítem.

En este caso no se eliminaron ítemes cubriendo todos los niveles críticos mencionados.

Se seleccionaron los 4 ítemes con el índice más alto de cada tema para integrar la “prueba FÁCIL”, los 4 ítemes con índices intermedios de cada tema para integrar la “prueba MEDIA” y los 4 ítemes con el índice más bajo para integrar la “prueba DIFÍCIL”. Los resultados de las “pruebas” corresponden a la misma aplicación en condiciones de admisión de alumnos al nivel medio superior, con todas sus características.

Se procesaron los 90 casos de cada una de las 10 versiones en cada una de las “pruebas” de 40 ítemes y por el instrumento original de 120 ítemes, incluyéndose los indicadores reportados por BIGSTEPS®; y se ordenó a los sujetos de mayor a menor frecuencia de aciertos en los 120 ítemes de cada una de las versiones, dividiéndolos en tres submuestras con respecto a la habilidad: “Alta”, “Media” y “Baja”.

Se procedió a correlacionar cada uno de los indicadores de cada una de las “pruebas” con los indicadores correspondientes del instrumento original (I = 120); los resultados se reportan en el anexo A, donde:

G = Grupo de habilidad (A: Alta, M: Media y B: Baja)

V = Versión del instrumento original (de 0 a 9)

PT = Puntaje de acierto de los sujetos

MED = Medición original en lógitos

ERR = Error de medición

IMS = Ajuste interno cuadrático medio

IZS = Ajuste interno estandarizado

OMS = Ajuste externo cuadrático medio

OZS = Ajuste externo estandarizado

PB = Correlación punto-biserial

## Resultados

Existe mayor relación en los puntajes (PT) cuando se elimina la dificultad relativa del instrumento.

- Podemos observar que al correlacionar a los 90 sujetos, sin importar la versión que respondan, el valor tiende a ser mayor.

En la prueba “fácil”, para el grupo de habilidad “alta” sólo dos versiones reportan una correlación mayor (A4 y A8), mientras que para los grupos de habilidad “media” y “baja” ninguna es mayor. En la prueba “media”,



para el grupo de habilidad “alta” sólo dos versiones reportan una correlación mayor (A4 y A9), para el grupo de habilidad “media” ninguna es mayor y para el grupo de habilidad “baja” sólo dos versiones reportan una correlación mayor (B5 y B9). En la prueba “difícil”, para el grupo de habilidad “alta” cuatro versiones reportan una correlación mayor (A0, A2, A4 y A8), para el grupo de habilidad “media” sólo dos versiones reportan una correlación mayor (M3 y M4) y para el grupo de habilidad “baja” sólo una versión reporta una correlación mayor (B5).

Existe mayor relación en la medición (MED) cuando se elimina la dificultad relativa del instrumento.

- Podemos observar que al correlacionar a los 90 sujetos, sin importar la versión que respondan, el valor tiende a ser mayor.

En la prueba “fácil”, para el grupo de habilidad “alta” sólo dos versiones reportan una correlación mayor (A4 y A8) y para los grupos de habilidad “media” y “baja” ninguna es mayor. En la prueba “media”, para el grupo de habilidad “alta” sólo tres versiones reportan una correlación mayor (A1, A4 y A9), para el grupo de habilidad “media” ninguna es mayor y para el grupo de habilidad “baja” sólo dos versiones reportan una correlación mayor (B5 y B9). En la prueba “difícil”, para el grupo de habilidad “alta” seis versiones reportan una correlación mayor (A0, A2, A3, A4, A7 y A8), para el grupo de habilidad “media” sólo tres versiones reportan una correlación mayor (M3, M4 y M5) y para el grupo de habilidad “baja” sólo dos versiones reportan una correlación mayor (B2 y B5).

Cuando se correlacionan los sujetos con independencia de la dificultad relativa de los instrumentos, existe un decremento en el valor de correlación de los indicadores relativos al puntaje (PT) y a la medición (MED) entre la prueba “fácil” y la “difícil”, afectando más al grupo de habilidad “baja” y menos al grupo de habilidad “alta”.

- Se observa que al correlacionar a los 90 sujetos, sin importar la versión que respondan, la correlación en el indicador de puntaje (PT) para el grupo de habilidad “alta” corresponde a: “fácil” = 0.842, “media” = 0.847 y

“difícil” = 0.824; para el grupo de habilidad “media” corresponde a: “fácil” = 0.784, “media” = 0.746 y “difícil” = 0.629; y para el grupo de habilidad “baja” corresponde a: “fácil” = 0.838, “media” = 0.730 y “difícil” = 0.563.

La correlación para el grupo de habilidad “alta” en el indicador relativo a la medición (MED) es: “fácil” = 0.858, “media” = 0.849 y “difícil” = 0.794; para el grupo de habilidad “media” es: “fácil” = 0.782, “media” = 0.729 y “difícil” = 0.570; y para el grupo de habilidad “baja” es: “fácil” = 0.820, “media” = 0.725 y “difícil” = 0.518.

Los únicos indicadores que resultan ser consistentemente significativos en la correlación entre las pruebas y el instrumento original, son el de puntaje (PT) y el de medición (MED) para el grupo de habilidad “alta”.

- Observamos que para el grupo de habilidad “media”, en ambos indicadores no es significativa la correlación en la versión M3 para la prueba “fácil”, en las versiones M4 y M9 para la prueba “media” y en las versiones M1 y M2 para la prueba “difícil”, sumándose la versión M9 en esta prueba para el indicador de puntaje (PT).

En el grupo de habilidad “baja” resultan significativas las correlaciones para los indicadores de puntaje (PT) y de medición (MED) en las diez versiones de las pruebas “fácil” y “media”, mientras que no son significativas para las versiones B3, B4, B6 y B8 en la prueba “difícil”.

Ninguna versión resulta ser significativa en todos los indicadores, para todas las pruebas y para las distintas habilidades.

- En la prueba “fácil” resultan significativas para el grupo de habilidad “alta” las versiones A1 y A9, para el grupo de habilidad “media” la versión M9 y para el grupo de habilidad “baja” las versiones B2, B4, B8 y B9; la versión 9 es la única común para esta prueba.

En la prueba “media” son significativas para el grupo de habilidad “alta” las versiones A2 y A8, para el grupo de habilidad “media” la versión M2 y para el grupo de habilidad “baja” las versiones B8 y B9.

En la prueba “difícil” es significativa para el grupo de habilidad “alta” la versión A8; para el grupo de habilidad “media” las versiones M0, M6 y M8; y para el grupo de habilidad “baja” las versiones B1, B2 y B5.

Los indicadores de desajuste no son estables en la significación resultante de las diferentes versiones, pruebas y habilidades.

- En la prueba “fácil” resultan significativas para el grupo de habilidad «alta» las versiones A1, A3 y A9; para el grupo de habilidad “media” las versiones M0, M1, M2, M3, M4, M7 y M9; para el grupo de habilidad “baja” las versiones B4, B8 y B9.

En la prueba “media” son significativas para el grupo de habilidad “alta” las versiones A2 y A8, para el grupo de habilidad “media” la versión M2 y para el grupo de habilidad “baja” las versiones B8 y B9.

En la prueba “difícil” se reportan significativas para el grupo de habilidad “alta” las versiones A0, A5 y A8; para el grupo de habilidad “media” las versiones M0, M1, M2, M6, M7 y M8; y para el grupo de habilidad “baja” las versiones B0, B1, B2, B4, B5 y B9.

Tomando las diferentes habilidades con independencia de las versiones, no se reportan diferencias superiores a 0.500.

- No obstante lo anterior, podemos mencionar que para los grupos de habilidad “media” y “baja” resultan superiores a 0.500 los indicadores de ajuste interno en la prueba “fácil” y los indicadores de ajuste externo en la prueba “difícil”; en ambos casos el ajuste cuadrático medio resulta ligeramente superior al ajuste estandarizado.

Las correlaciones en los indicadores no demuestran que se mantenga una jerarquía similar entre los lugares ocupados en el instrumento completo y las pruebas “fácil”, “media” y “difícil”.

- Ninguno de los indicadores, al margen de la dificultad de la prueba o de la habilidad de los sujetos, llega a reportar valores iguales o mayores que 0.900 ni que 0.800 de correlación en las 10 versiones.

En correlaciones de 0.700 o mayores, sólo el grupo de habilidad “alta” en la prueba “difícil” llega a las 10 versiones para el indicador del puntaje (PT).

Para las correlaciones de 0.600 o mayores, se incluye la medición (MED) en el grupo de habilidad “alta” en las pruebas “fácil” y “difícil”, que llegan a las 10 versiones.

En correlaciones de 0.500 o mayores, el grupo de habilidad “alta” añade el puntaje (PT) en la prueba “fácil” y “media”, así como la medición (MED) en la prueba “media”, mientras que el grupo de habilidad “baja” incluye la puntuación (PT) y la medición (MED) en la prueba “fácil”, así como el error (ERR) para la prueba “media” en las 10 versiones.

Para las correlaciones de 0.400 o mayores, el grupo de habilidad “baja” añade los valores de puntuación (PT) y medición (MED) para la prueba “media” en las 10 versiones.

Para correlaciones mayores o iguales que 0.300, los grupos de habilidad “media” y “baja” añaden los valores de ajustes externos cuadrático medio (OMS) y estandarizado (OZS) para la prueba “difícil”, y el grupo de habilidad “baja” añade los valores de ajustes internos medio cuadrático (IMS) y estandarizado (IZS) para la prueba “fácil” en las 10 versiones.

En la correlaciones de 0.200 o mayores, los grupos de habilidad “media” y “baja” añaden los valores de puntuación (PT) y medición (MED) para la prueba “difícil” y los valores de punto-biserial (PB) para la prueba “fácil”, y el grupo de habilidad “media” añade los valores de ajustes internos me-

dio cuadrático (IMS) y estandarizado (IZS), además del valor de puntaje (PT) para la prueba “media” en la 10 versiones.

Para las correlaciones superiores o iguales a 0.100, el grupo de habilidad “alta” añade para las pruebas “fácil” y “difícil” los valores de ajuste interno medio cuadrático (IMS) y para la prueba “media” el valor de error (ERR); el grupo de habilidad “media” incluye los valores de medición (MED) para la prueba “media” y de ajuste interno estandarizado (IZS) para la prueba “difícil”; el grupo de habilidad “baja” añade los valores de error (ERR) para las pruebas “fácil” y “difícil”, así como los valores de ajuste interno estandarizado (IZS) y de punto-biserial (PB) para la prueba “difícil” en las 10 versiones.

En las correlaciones de 0.000 o mayores, el grupo de habilidad “alta” incluye los valores de ajuste externo medio cuadrático (OMS) para las pruebas “fácil” y “difícil”, así como los valores de ajuste interno estandarizado (IZS) y de punto-biserial (PB) para la prueba “fácil”; el grupo de habilidad “media” incluye los valores de ajustes externos medio cuadrático (OMS) y estandarizado (OZS) para la prueba “fácil”, y los valores de ajuste interno medio cuadrático (IMS) y de punto-biserial (PB) para la prueba “difícil”; el grupo de habilidad “baja” incluye el valor de ajuste interno medio cuadrático (IMS) para las pruebas “media” y “difícil”, así como los valores de ajuste externo medio cuadrático (OMS) y de punto-biserial (PB) para la prueba “media” en las 10 versiones.

## **Conclusiones**

1. Es importante contar con un valor central estable que sea independiente de la dificultad relativa del instrumento así como de las variaciones relativas con respecto a los temas que lo integran. Sin embargo, esto no exime al constructor de instrumentos de medición del cuidado que debe tenerse al elegir la mínima diferencia entre las dificultades de los ítemes entre las versiones que desee que sean equivalentes, ya que las oscilaciones en los valores distantes del punto central establecerán diferencias en el lugar que ocupen los sujetos evaluados.

2. No basta con que las pruebas contengan ítemes tomados siguiendo una misma tabla de especificaciones para integrarlas, sino que los ítemes considerados equivalentes deben tener dificultades muy próximas entre sí para que al formar un instrumento éste sea relativamente equivalente.

3. Si bien en promedio (ver Anexo B) el modelo logístico mantiene invarianza en los indicadores, la dispersión con respecto al valor central es variable, lo que hace cambiar el lugar que ocupa una persona en función de la dificultad de la prueba que se le aplique.

4. Dado que esta metodología no se ocupa de los índices de discriminación, la base de datos para realizar el análisis debe evitar sesgos con respecto a las calificaciones de la muestra de sujetos, así como a la calidad de los ítemes que integran el instrumento. Un sujeto con un alto nivel de preguntas sin respuesta se verá favorecido si no se toman las faltas de respuesta como errores al contar con una mayor cantidad de tiempo para resolver las preguntas que responde.

5. Si se encuentran sujetos que no tengan la habilidad necesaria o que se excedan de la habilidad esperada dentro de su grupo de referencia, se propiciará un sesgo que puede repercutir en los valores distantes del punto central del modelo.

6. Si en la muestra se encuentran sujetos cuyas respuestas difieran de manera suficiente de las esperadas para su grupo de referencia, el incluirlo puede propiciar un sesgo en los resultados distales con respecto al punto central del modelo.

En muchos casos nos interesarán los valores centrales de respuesta, pero en situaciones de toma de decisiones en límites de corte sobre la calidad de respuesta de los sujetos, es deseable contar con valores estables lejanos al punto central, los cuales oscilan en el modelo con respecto al lugar que ocupa el sujeto de acuerdo con el instrumento aplicado.

**ANEXO A**

**Comparación entre rechazados y aceptados**

Ver	Gpo	Med	Var	N	Var/N		Dif-M	EPM	gl	t	
V0	R	40.800	297.853	20	14.893	RA	6.443	4.167	88	1.546	
	A	47.243	172.853	70	2.469	RT	5.011	4.144	108	1.209	
	T	45.811	204.852	90	2.276	AT	1.432	2.178	158	0.657	
V1	R	37.682	129.465	22	5.885	RA	15.112	2.872	88	5.262	*
	A	52.794	160.614	68	2.362	RT	11.418	2.836	110	4.026	*
	T	49.100	194.113	90	2.157	AT	3.694	2.126	156	1.738	
V2	R	33.882	207.860	17	12.227	RA	11.720	3.805	88	3.080	*
	A	45.603	164.187	73	2.249	RT	9.507	3.789	105	2.509	*
	T	43.389	191.476	90	2.128	AT	2.214	2.092	161	1.058	
V3	R	34.880	138.777	25	5.551	RA	10.320	2.798	88	3.689	*
	A	45.200	147.881	65	2.275	RT	7.453	2.718	113	2.742	*
	T	42.333	165.371	90	1.837	AT	2.867	2.028	153	1.414	
V4	R	31.500	47.167	10	4.717	RA	15.800	2.814	88	5.615	*
	A	47.300	256.213	80	3.203	RT	14.044	2.752	98	5.103	*
	T	45.544	257.127	90	2.857	AT	1.756	2.462	168	0.713	
V5	R	45.000	322.000	6	53.667	RA	16.440	7.509	88	2.189	*
	A	61.440	228.394	84	2.719	RT	15.344	7.512	94	2.043	*
	T	60.344	248.094	90	2.757	AT	1.096	2.340	172	0.468	
V6	R	36.913	90.628	23	3.940	RA	9.744	2.554	88	3.815	*
	A	46.657	173.138	67	2.584	RT	7.254	2.412	111	3.007	*
	T	44.167	169.062	90	1.878	AT	2.490	2.112	155	1.179	
V7	R	44.286	98.571	7	14.082	RA	12.389	4.051	88	3.058	*
	A	56.675	193.222	83	2.328	RT	11.425	4.032	95	2.834	*
	T	55.711	195.803	90	2.176	AT	0.964	2.122	171	0.454	
V8	R	27.353	47.868	17	2.816	RA	13.387	2.287	88	5.854	*
	A	40.740	176.251	73	2.414	RT	10.858	2.192	105	4.954	*
	T	38.211	178.955	90	1.988	AT	2.529	2.098	161	1.205	
V9	R	36.406	192.249	32	6.008	RA	8.783	3.038	88	2.891	*
	A	45.190	186.788	58	3.220	RT	5.660	2.877	120	1.967	*
	T	42.067	204.467	90	2.272	AT	3.123	2.344	146	1.333	
Tot	R	36.128	165.573	179	0.925	RA	13.156	1.111	898	11.837	*
	A	49.284	223.623	721	0.310	RT	10.539	1.091	1077	9.657	*
	T	46.668	239.488	900	0.266	AT	2.617	0.759	1619	3.447	*

**ANEXO B**

**Comparación entre procedimientos  
(Promedios por versión)**

Ver	Procedimiento "A" Grupo completo					Procedimiento "B" Grupo depurado				
	I	S	SE	Dif	Dis	I	S	SE	Dif	Dis
V0	120	90	0.985	0.382	0.247	120	90	0.994	0.394	0.375
	78	90	0.987	0.390	0.325	107	70	0.993	0.408	0.404
V1	120	90	0.977	0.409	0.238	120	90	0.995	0.439	0.367
	81	90	0.978	0.447	0.318	102	69	0.995	0.460	0.409
V2	120	90	1.000	0.362	0.245	120	90	1.000	0.380	0.357
	79	90	1.000	0.389	0.316	101	73	1.000	0.377	0.386
V3	120	90	0.966	0.353	0.215	120	90	0.991	0.377	0.347
	68	90	0.961	0.368	0.306	99	65	0.990	0.397	0.397
V4	120	90	1.000	0.380	0.280	120	90	1.000	0.394	0.397
	88	90	1.000	0.402	0.345	104	80	1.000	0.408	0.433
V5	120	90	0.984	0.503	0.263	120	90	0.991	0.512	0.383
	82	90	0.984	0.526	0.346	104	84	0.990	0.522	0.424
V6	120	90	0.994	0.368	0.226	120	90	0.995	0.389	0.381
	71	90	0.994	0.396	0.321	105	67	0.995	0.401	0.418
V7	120	90	0.992	0.468	0.241	120	90	0.997	0.476	0.380
	72	90	0.992	0.485	0.328	109	83	0.997	0.476	0.403
V8	120	90	0.979	0.318	0.237	120	90	0.988	0.339	0.379
	69	90	0.980	0.340	0.330	106	73	0.988	0.347	0.409
V9	120	90	0.982	0.351	0.244	120	90	0.995	0.377	0.371
	74	90	0.983	0.351	0.328	101	58	0.995	0.376	0.408
Pro	120	90	0.986	0.389	0.244	120	90	0.995	0.408	0.374
	76	90	0.986	0.409	0.326	104	72	0.994	0.417	0.409

Promedios		Sujetos	
		Totales	Depurados
Ítemes	Totales	90-120	Dep-120
	Depurados	90-Dep	Dep-Dep

- La primera línea incluye los 120 ítemes.
- La segunda línea incluye sólo los ítemes aceptados.
- El procedimiento "A" incluye los 90 sujetos.
- El procedimiento "B" incluye sólo los sujetos aceptados.



ANEXO C

**CORRELACIONES**  
**Pruebas: Fácil, Media y Difícil contra Total**  
**Habilidades: Alta, Media, Baja**

IMS	IZS	OMS	OZS	PB	PT	MED	ERR	IMS	IZS	OMS	OZS	PB	PT	MED	ERR	IMS	IZS	OMS	OZS	PB
0.27	0.22	0.52	0.41	0.35	0.63	0.63	0.33	0.28	0.33	0.37	0.32	0.35	0.85	0.83	-0.25	0.11	-0.00	0.04	-0.03	-0.02
0.40	0.45	0.43	0.55	0.40	0.84	0.86	0.91	0.22	0.13	0.10	0.06	0.21	0.74	0.73	-0.26	0.47	0.36	0.51	0.42	0.39
0.40	0.34	0.57	0.52	0.42	0.79	0.80	0.82	0.47	0.51	0.45	0.47	0.44	0.85	0.84	-0.10	0.11	0.01	0.30	0.18	0.04
0.54	0.50	0.64	0.60	0.53	0.84	0.84	0.13	0.42	0.39	0.32	0.29	0.41	0.82	0.80	0.19	0.26	0.07	0.41	0.16	0.20
0.31	0.41	0.57	0.59	0.32	0.91	0.91	0.83	0.19	-0.01	0.13	-0.00	0.14	0.85	0.83	-0.35	0.34	0.22	0.25	0.20	0.14
0.10	0.20	0.09	0.50	0.05	0.74	0.79	0.89	0.54	0.53	0.34	0.39	0.53	0.77	0.76	-0.09	0.58	0.62	0.36	0.42	0.48
0.43	0.27	0.41	0.32	0.44	0.74	0.74	0.28	-0.03	-0.09	-0.06	-0.10	-0.06	0.72	0.70	-0.11	0.28	0.23	0.34	0.30	0.20
0.14	0.11	0.51	0.44	0.17	0.56	0.57	0.47	0.23	0.21	0.23	0.23	0.18	0.83	0.83	-0.26	0.41	0.26	0.31	0.25	0.31
0.11	0.03	0.09	-0.00	0.10	0.76	0.76	0.63	0.49	0.39	0.44	0.38	0.46	0.90	0.88	0.54	0.48	0.40	0.57	0.51	0.44
0.38	0.41	0.35	0.43	0.39	0.92	0.92	0.68	0.41	0.36	0.36	0.33	0.39	0.78	0.76	-0.26	0.37	0.35	0.50	0.44	0.36
0.70	0.71	0.60	0.63	0.64	0.39	0.39	0.28	0.25	0.20	0.12	0.11	0.24	0.48	0.48	0.36	0.45	0.48	0.57	0.64	0.43
0.77	0.71	0.70	0.64	0.75	0.59	0.59	0.29	-0.15	-0.12	-0.12	-0.07	-0.12	0.27	0.28	0.38	0.41	0.58	0.43	0.61	0.29
0.72	0.71	0.46	0.47	0.68	0.41	0.41	0.39	0.53	0.41	0.54	0.45	0.51	0.34	0.34	-0.05	0.46	0.52	0.76	0.73	0.45
0.73	0.72	0.59	0.60	0.71	0.44	0.44	0.50	0.13	0.20	0.03	0.12	0.12	0.73	0.69	0.42	0.06	0.11	0.30	0.30	0.03
0.57	0.57	0.45	0.48	0.53	0.24	0.24	0.18	0.30	0.28	0.32	0.33	0.26	0.63	0.62	0.46	0.47	0.33	0.58	0.61	0.46
0.47	0.41	0.04	0.07	0.44	0.52	0.52		0.37	0.39	0.14	0.18	0.38	0.60	0.60		0.30	0.29	0.96	0.90	0.10
0.28	0.21	0.15	0.10	0.28	0.52	0.50	0.34	0.34	0.39	0.23	0.30	0.34	0.53	0.53	0.40	0.41	0.38	0.63	0.60	0.44
0.52	0.53	0.54	0.60	0.46	0.48	0.46	-0.08	0.18	0.26	0.09	0.18	0.16	0.53	0.53	0.59	0.39	0.36	0.52	0.56	0.33
0.34	0.31	0.60	0.58	0.26	0.21	0.20	0.22	0.38	0.31	0.20	0.12	0.40	0.44	0.49	0.52	0.47	0.50	0.67	0.71	0.48
0.69	0.69	0.62	0.60	0.71	0.30	0.29	0.05	0.30	0.41	0.16	0.27	0.30	0.33	0.36	0.23	0.35	0.31	0.52	0.47	0.35
0.43	0.41	0.09	0.15	0.29	0.55	0.55	0.56	0.39	0.35	0.42	0.35	0.43	0.43	0.38	0.24	0.47	0.56	0.66	0.78	0.49
0.54	0.44	0.24	0.13	0.43	0.57	0.60	0.61	0.07	-0.04	0.17	0.13	0.14	0.39	0.44	0.48	0.50	0.52	0.69	0.78	0.47
0.82	0.80	0.48	0.57	0.78	0.60	0.61	0.65	0.50	0.28	0.36	0.22	0.49	0.48	0.63	0.75	0.37	0.42	0.86	0.81	0.43
0.61	0.52	0.46	0.35	0.56	0.58	0.61	0.66	0.02	0.05	0.01	-0.04	0.06	0.24	0.22	0.12	0.30	0.31	0.34	0.44	0.25
0.64	0.62	0.47	0.46	0.63	0.44	0.50	0.59	0.33	0.40	0.25	0.37	0.38	0.25	0.28	0.31	0.42	0.41	0.56	0.65	0.33
0.46	0.48	-0.12	-0.11	0.28	0.76	0.76	0.66	0.19	0.17	0.01	0.10	0.16	0.58	0.60	0.64	0.49	0.55	0.88	0.93	0.49
0.44	0.38	0.21	0.23	0.36	0.62	0.66	0.69	0.20	0.09	0.08	0.09	0.16	0.20	0.20	0.18	0.05	0.13	0.31	0.39	0.12
0.33	0.33	0.15	0.16	0.23	0.60	0.60	0.62	0.53	0.52	0.40	0.34	0.57	0.48	0.45	0.26	0.19	0.24	0.47	0.55	0.13
0.50	0.54	0.45	0.53	0.38	0.65	0.65	0.64	0.46	0.46	0.41	0.39	0.44	0.24	0.23	0.16	0.27	0.29	0.32	0.40	0.23
0.80	0.81	0.66	0.70	0.80	0.75	0.78	0.79	0.45	0.36	0.40	0.38	0.46	0.42	0.40	0.34	0.45	0.47	0.58	0.63	0.40
0.28	0.28	0.32	0.44	0.26	0.85	0.85	0.77	0.30	0.24	0.20	0.19	0.27	0.82	0.79	-0.08	0.38	0.25	0.38	0.28	0.39
0.61	0.57	0.37	0.47	0.54	0.75	0.73	0.44	0.25	0.26	0.15	0.19	0.31	0.63	0.57	0.46	0.38	0.40	0.78	0.65	0.41
0.58	0.51	0.25	0.25	0.52	0.73	0.72	0.67	0.30	0.25	0.17	0.20	0.38	0.56	0.52	0.52	0.36	0.41	0.73	0.71	0.37

Los grupos están formados por n = 30 y las pruebas por i = 40.

Valores extremos y promedio en los indicadores  
(Pruebas fácil- media-difícil en sus 10 versines)

ANEXO D

	Fácil							Media							Difícil									
	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB
	Versión 0							Versión 0							Versión 0									
<b>Min</b>	9	-1.28	0.32	0.86	-1.45	0.78	-1.50	-0.26	4	-2.21	0.32	0.94	-1.30	0.92	-1.29	-0.38	3	-2.56	0.32	0.87	-1.68	0.81	-1.68	-0.57
<b>Máx</b>	34	1.79	0.45	1.15	2.16	1.36	2.20	0.49	29	0.98	0.53	1.07	1.25	1.10	1.27	0.30	24	0.41	0.60	1.13	0.88	1.63	1.40	0.46
<b>Var</b>	25	3.07	0.13	0.29	3.61	0.58	3.70	0.75	25	3.19	0.21	0.13	2.55	0.18	2.56	0.68	21	2.97	0.28	0.26	2.56	0.82	3.08	1.03
<b>Pro</b>	21	0.10	0.34	1.00	0.00	1.00	-0.01	0.14	15	-0.58	0.35	1.00	0.00	1.00	0.00	-0.03	10	-1.18	0.39	1.00	-0.03	1.01	-0.02	0.05
	Versión 1							Versión 1							Versión 1									
<b>Min</b>	9	-1.31	0.33	0.82	-1.73	0.70	-1.73	-0.26	5	-1.97	0.32	0.92	-1.78	0.87	-1.79	-0.38	3	-2.59	0.32	0.88	-1.69	0.77	-1.65	-0.19
<b>Máx</b>	37	2.63	0.61	1.22	2.41	1.37	2.46	0.56	34	1.76	0.48	1.10	1.68	1.16	1.67	0.35	23	0.31	0.60	1.12	1.69	1.37	1.85	0.45
<b>Var</b>	28	3.94	0.28	0.40	4.14	0.67	4.19	0.82	29	3.73	0.16	0.18	3.46	0.29	3.46	0.73	20	2.90	0.28	0.24	3.38	0.60	3.50	0.64
<b>Pro</b>	22	0.30	0.35	1.00	-0.04	1.00	-0.06	0.19	15	-0.50	0.34	1.00	-0.02	1.00	-0.02	0.04	11	-1.04	0.37	1.00	-0.03	1.00	-0.03	0.10
	Versión 2							Versión 2							Versión 2									
<b>Min</b>	6	-1.85	0.33	0.79	-2.74	0.73	-2.50	-0.16	4	-2.22	0.32	0.90	-1.42	0.89	-1.48	-0.48	0	-4.44	0.33	0.87	-2.27	0.78	-1.66	-0.22
<b>Máx</b>	30	1.18	0.46	1.20	2.16	1.34	2.33	0.58	29	0.98	0.53	1.11	1.65	1.15	1.66	0.47	26	0.71	1.42	1.11	0.59	2.57	2.72	0.45
<b>Var</b>	24	3.03	0.13	0.41	4.90	0.61	4.83	0.74	25	3.20	0.21	0.21	3.07	0.26	3.14	0.95	26	5.15	1.09	0.24	2.86	1.79	4.38	0.67
<b>Pro</b>	20	0.02	0.34	1.00	-0.07	1.00	-0.07	0.21	14	-0.70	0.35	1.00	-0.03	1.00	-0.03	0.00	9	-1.31	0.41	1.00	-0.04	1.01	-0.03	0.08
	Versión 3							Versión 3							Versión 3									
<b>Min</b>	7	-1.62	0.32	0.82	-1.94	0.81	-1.86	-0.27	4	-2.22	0.32	0.91	-1.85	0.86	-1.86	-0.28	3	-2.57	0.32	0.91	-1.50	0.78	-1.46	-0.32
<b>Máx</b>	33	1.62	0.42	1.22	2.59	1.24	2.61	0.55	28	0.86	0.53	1.07	1.76	1.13	1.78	0.39	19	-0.11	0.60	1.11	0.61	1.35	0.99	0.37
<b>Var</b>	26	3.24	0.10	0.40	4.53	0.43	4.47	0.82	24	3.08	0.21	0.16	3.61	0.27	3.64	0.67	16	2.46	0.28	0.20	2.11	0.57	2.45	0.69
<b>Pro</b>	20	-0.04	0.34	1.00	-0.03	1.00	-0.02	0.17	14	-0.69	0.35	1.00	-0.01	1.00	-0.01	0.01	9	-1.36	0.41	1.00	-0.06	1.02	-0.03	0.06
	Versión 4							Versión 4							Versión 4									
<b>Min</b>	4	-2.32	0.33	0.81	-2.58	0.76	-2.38	-0.12	4	-2.24	0.32	0.91	-1.59	0.82	-1.65	-0.34	3	-2.58	0.32	0.82	-2.72	0.81	-2.72	-0.30
<b>Máx</b>	36	2.30	0.54	1.16	1.99	1.43	1.87	0.57	32	1.41	0.53	1.11	1.10	1.19	1.10	0.37	25	0.53	0.60	1.14	1.00	1.40	1.46	0.58
<b>Var</b>	32	4.62	0.21	0.35	4.57	0.67	4.25	0.69	28	3.65	0.21	0.20	2.69	0.37	2.75	0.71	22	3.11	0.28	0.32	3.72	0.59	4.18	0.88
<b>Pro</b>	21	0.14	0.35	1.00	-0.02	1.00	-0.02	0.19	15	-0.59	0.35	1.00	-0.04	1.00	-0.02	0.06	10	-1.29	0.40	1.00	-0.11	1.02	-0.06	0.08

180

Diferencias entre los valores de las pruebas

ANEXO E

Valores extremos y promedio en los indicadores  
(Pruebas fácil-media-difícil en sus 10 versiones)

	Fácil								Media								Difícil							
	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB
	Versión 5								Versión 5								Versión 5							
<b>Min</b>	14	-0.66	0.33	0.76	-3.21	0.68	-2.84	-0.25	7	-1.61	0.32	0.86	-2.26	0.81	-2.15	-0.25	4	-2.45	0.34	0.82	-1.85	0.70	-1.26	-0.16
<b>Máx</b>	39	3.81	1.02	1.16	1.87	2.67	2.27	0.65	36	2.27	0.53	1.16	1.18	1.22	1.23	0.50	25	0.50	0.53	1.22	1.33	4.00	4.35	0.57
<b>Var</b>	25	4.47	0.69	0.40	5.08	1.99	5.11	0.90	29	3.88	0.21	0.30	3.44	0.41	3.38	0.75	21	2.95	0.19	0.40	3.18	3.30	5.61	0.73
<b>Pro</b>	27	0.96	0.39	1.00	-0.03	1.01	-0.04	0.18	20	0.01	0.34	1.00	-0.04	1.00	-0.07	0.13	13	-0.97	0.37	1.00	-0.07	1.05	-0.03	0.23
	Versión 6								Versión 6								Versión 6							
<b>Min</b>	11	-0.99	0.32	0.88	-2.07	0.85	-2.06	-0.29	3	-2.56	0.32	0.85	-3.16	0.78	-3.14	-0.34	3	-2.56	0.32	0.87	-1.48	0.76	-1.59	-0.37
<b>Máx</b>	35	1.98	0.48	1.12	2.24	1.19	2.18	0.47	26	0.63	0.60	1.12	1.21	1.21	1.26	0.56	24	0.42	0.60	1.11	0.78	1.31	1.09	0.49
<b>Var</b>	24	2.97	0.16	0.24	4.31	0.34	4.24	0.76	23	3.19	0.28	0.27	4.37	0.43	4.40	0.90	21	2.98	0.28	0.24	2.26	0.55	2.68	0.86
<b>Pro</b>	20	0.03	0.34	1.00	-0.02	1.00	-0.03	0.08	14	-0.65	0.35	1.00	-0.05	1.00	-0.03	0.07	10	-1.26	0.40	1.00	-0.03	1.01	-0.02	0.06
	Versión 7								Versión 7								Versión 7							
<b>Min</b>	11	-1.01	0.32	0.86	-1.89	0.75	-1.79	-0.30	9	-1.27	0.32	0.90	-1.74	0.89	-1.75	-0.39	4	-2.26	0.33	0.86	-1.54	0.78	-1.57	-0.20
<b>Máx</b>	37	2.60	0.60	1.22	2.69	1.44	2.73	0.47	30	1.13	0.38	1.13	1.91	1.26	1.94	0.40	28	0.98	0.53	1.15	1.65	1.40	2.05	0.46
<b>Var</b>	26	3.61	0.28	0.36	4.58	0.69	4.52	0.77	21	2.40	0.06	0.23	3.65	0.37	3.69	0.79	24	3.24	0.20	0.29	3.19	0.62	3.62	0.66
<b>Pro</b>	25	0.61	0.36	1.00	-0.06	1.00	-0.04	0.15	18	-0.18	0.33	1.00	0.00	1.00	-0.01	0.09	12	-0.87	0.37	1.00	-0.06	1.01	-0.03	0.15
	Versión 8								Versión 8								Versión 8							
<b>Min</b>	3	-2.72	0.33	0.82	-1.87	0.79	-1.13	-0.12	4	-2.23	0.32	0.90	-0.99	0.83	-1.02	-0.39	3	-2.56	0.32	0.88	-1.18	0.74	-1.26	-0.31
<b>Máx</b>	33	1.65	0.64	1.18	2.60	1.58	3.55	0.47	26	0.63	0.53	1.12	1.05	1.19	1.10	0.45	23	0.31	0.60	1.10	0.55	1.41	0.93	0.51
<b>Var</b>	30	4.37	0.31	0.36	4.47	0.79	4.68	0.59	22	2.86	0.21	0.22	2.04	0.36	2.12	0.84	20	2.87	0.28	0.22	1.73	0.67	2.19	0.82
<b>Pro</b>	17	-0.28	0.35	1.00	0.01	0.99	-0.06	0.20	12	-0.87	0.36	1.00	-0.02	1.00	-0.03	0.05	8	-1.45	0.42	1.00	-0.02	1.00	-0.02	0.05

Diferencias entre los valores de las pruebas

Rango de las diferencias entre valores del instrumento total y cada una de las pruebas (Por pruebas y habilidad del grupo)

G	V	PT	FÁCIL						MEDIA						DIFÍCIL										
			MED	ERR	LMS	LZS	O.MS	O.ZS	PB	PT	MED	ERR	LMS	LZS	O.MS	O.ZS	PB	PT	MED	ERR	LMS	LZS	O.MS	O.ZS	PB
A	0	25	1.44	0.12	0.24	4.06	0.36	3.19	0.54	27	1.11	0.09	0.17	3.50	0.22	3.40	0.53	21	0.97	0.11	0.28	4.92	0.42	5.54	0.75
M	0	12	1.00	0.02	0.24	2.80	0.27	2.86	0.59	13	0.82	0.05	0.22	2.92	0.25	3.04	0.72	12	0.98	0.10	0.19	2.06	0.26	2.00	0.60
B	0	9	0.88	0.06	0.21	2.36	0.42	2.79	0.50	11	1.59	0.19	0.16	1.66	0.39	2.69	0.61	16	1.62	0.24	0.21	1.50	0.61	2.22	0.77
A	1	31	1.03	0.24	0.27	2.56	0.54	2.88	0.56	24	0.88	0.09	0.21	4.21	0.27	4.21	0.57	29	0.95	0.08	0.18	3.43	0.29	3.17	0.40
M	1	14	1.45	0.07	0.25	4.20	0.37	4.16	0.47	11	0.66	0.05	0.42	6.25	0.65	6.31	0.94	19	1.73	0.13	0.28	4.96	0.42	4.75	0.60
B	1	14	1.16	0.06	0.30	2.99	0.71	4.62	0.68	18	1.31	0.14	0.26	2.60	0.56	3.37	0.76	19	1.52	0.22	0.21	2.11	0.43	2.48	0.44
A	2	31	1.14	0.03	0.30	3.71	0.44	3.29	0.57	27	1.02	0.02	0.15	3.17	0.21	3.52	0.64	21	1.22	0.10	0.25	4.46	0.27	3.16	0.54
M	2	14	1.19	0.03	0.23	3.80	0.71	6.28	0.55	15	1.24	0.09	0.26	3.32	0.42	4.00	0.58	13	1.09	0.13	0.24	3.54	0.63	3.13	0.43
B	2	18	1.27	0.11	0.21	1.89	1.06	4.02	0.43	19	1.20	0.13	0.27	2.53	1.09	4.80	0.67	21	2.14	0.96	0.34	2.60	0.91	2.58	0.63
A	3	26	1.14	0.11	0.20	2.89	0.25	2.56	0.48	22	0.69	0.04	0.21	3.76	0.25	3.80	0.58	24	1.12	0.16	0.20	3.69	0.30	3.93	0.53
M	3	18	1.47	0.03	0.23	3.00	0.36	3.66	0.47	13	1.58	0.15	0.28	3.33	0.38	3.73	0.69	7	1.57	0.25	0.36	4.21	0.51	3.55	0.74
B	3	15	1.62	0.08	0.25	3.03	0.38	3.48	0.56	15	1.49	0.17	0.26	2.45	0.38	2.84	0.66	17	1.41	0.23	0.23	2.12	0.61	2.21	0.56
A	4	30	1.15	0.19	0.24	3.08	0.53	2.70	0.48	28	0.85	0.07	0.23	4.22	0.33	4.62	0.53	29	1.19	0.15	0.24	3.63	0.47	3.77	0.74
M	4	16	1.31	0.02	0.32	3.31	0.48	3.78	0.59	15	1.21	0.06	0.32	3.79	0.49	4.27	0.73	11	1.28	0.19	0.26	3.22	0.58	2.84	0.59
B	4	14	1.98	0.19	0.28	2.37	0.68	3.56	0.50	16	1.27	0.16	0.35	2.54	0.80	3.77	0.66	17	1.57	0.23	0.26	2.12	0.63	2.70	0.63
A	5	25	1.95	0.64	0.23	2.48	2.02	2.25	0.75	21	1.52	0.16	0.23	2.66	0.49	2.92	0.54	23	0.97	0.03	0.25	2.55	0.51	2.56	0.43
M	5	14	1.30	0.12	0.26	3.95	1.71	8.04	0.41	14	1.20	0.03	0.23	3.50	1.48	7.24	0.48	17	1.06	0.09	0.28	3.63	1.91	2.85	0.43
B	5	15	0.72	0.04	0.38	4.71	1.17	6.17	0.75	17	0.97	0.09	0.31	3.27	1.10	4.82	0.64	18	1.30	0.16	0.27	2.48	1.57	1.77	0.44
A	6	25	1.46	0.14	0.21	4.42	0.33	4.25	0.66	30	1.16	0.04	0.27	5.88	0.30	5.80	0.72	25	1.28	0.11	0.26	4.96	0.29	4.62	0.63
M	6	15	0.91	0.03	0.24	4.18	0.36	4.54	0.63	13	1.11	0.06	0.25	3.05	0.40	3.28	0.66	15	1.60	0.24	0.23	2.79	0.30	2.52	0.57
B	6	15	0.97	0.03	0.21	2.39	0.43	3.73	0.68	13	1.58	0.24	0.25	2.12	0.67	3.07	0.77	16	1.58	0.24	0.23	2.01	0.47	2.45	0.63
A	7	22	1.75	0.25	0.26	3.33	0.49	3.09	0.53	26	1.02	0.05	0.28	4.62	0.31	4.58	0.71	17	1.28	0.03	0.27	4.01	0.34	3.91	0.54
M	7	14	1.14	0.04	0.26	3.28	0.38	3.30	0.52	17	1.31	0.05	0.31	4.63	0.43	5.29	0.71	14	1.57	0.11	0.34	5.04	0.52	4.58	0.70
B	7	13	1.20	0.04	0.34	4.56	0.52	4.62	0.77	17	1.28	0.05	0.22	2.66	0.37	3.73	0.53	18	1.35	0.13	0.22	2.27	0.49	2.50	0.53
A	8	25	1.12	0.10	0.24	3.98	0.25	2.42	0.43	31	1.06	0.04	0.15	2.61	0.18	2.00	0.42	26	0.93	0.16	0.17	2.73	0.21	1.90	0.56
M	8	11	0.87	0.03	0.26	3.11	0.61	3.48	0.51	13	0.90	0.07	0.19	1.88	0.42	3.20	0.69	11	1.58	0.25	0.17	1.85	0.30	2.31	0.58
B	8	12	1.55	0.25	0.31	2.17	0.36	1.89	0.54	11	1.21	0.17	0.18	1.45	0.35	2.24	0.57	17	1.58	0.23	0.22	1.50	0.65	2.02	0.90
A	9	30	1.16	0.09	0.28	4.01	0.36	3.79	0.56	21	0.92	0.04	0.24	3.97	0.31	4.35	0.53	30	1.10	0.11	0.28	3.36	0.31	3.03	0.81
M	9	16	1.13	0.03	0.26	2.71	0.33	3.23	0.60	13	1.16	0.08	0.24	2.72	0.36	2.94	0.56	12	1.49	0.25	0.26	3.01	0.37	2.89	0.56
B	9	16	0.93	0.06	0.20	1.70	0.37	2.56	0.48	13	1.85	0.36	0.23	2.30	0.55	2.96	0.53	16	1.52	0.24	0.23	2.24	0.45	2.27	0.67

**Promedio de las diferencias entre valores del instrumento total y cada una de las pruebas  
(Por prueba y total de sujetos)**

G	V	FÁCIL							MEDIA							DIFÍCIL									
		PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB
T	0	25	-0.65	-0.14	0.00	-0.09	0.01	-0.02	0.09	31	0.03	-0.15	0.00	-0.09	0.01	-0.04	0.26	36	0.63	-0.19	0.00	-0.06	0.00	-0.01	0.19
T	1	27	-0.71	-0.15	0.00	-0.10	0.01	0.03	0.07	34	0.08	-0.14	0.00	-0.10	0.02	0.00	0.22	38	0.62	-0.17	0.00	-0.10	0.01	0.00	0.16
T	2	23	-0.66	-0.14	0.00	-0.01	0.01	0.04	0.05	30	0.06	-0.15	0.00	-0.05	0.02	0.00	0.25	34	0.66	-0.21	0.00	-0.04	0.00	0.00	0.17
T	3	23	-0.65	-0.14	0.00	-0.07	0.01	0.01	0.08	28	0.00	-0.15	0.00	-0.09	0.01	0.01	0.23	33	0.67	-0.20	0.00	-0.04	0.00	0.02	0.19
T	4	24	-0.72	-0.14	0.00	-0.07	0.01	0.03	0.08	31	0.02	-0.15	0.00	-0.06	0.01	0.03	0.21	36	0.72	-0.20	0.00	0.01	-0.01	0.07	0.20
T	5	33	-0.97	-0.18	0.00	-0.03	0.02	0.11	0.18	40	-0.03	-0.14	0.00	-0.02	0.03	0.14	0.22	47	0.95	-0.17	0.00	0.01	-0.02	0.10	0.12
T	6	24	-0.64	-0.13	0.00	-0.06	0.01	-0.01	0.15	30	0.03	-0.15	0.00	-0.03	0.00	0.00	0.16	35	0.65	-0.20	0.00	-0.05	0.00	-0.02	0.17
T	7	31	-0.76	-0.16	0.00	0.03	0.00	0.04	0.13	37	0.03	-0.13	0.00	-0.03	0.00	0.01	0.19	43	0.72	-0.17	0.00	0.03	-0.01	0.03	0.13
T	8	21	-0.58	-0.14	0.00	-0.09	0.01	0.02	0.03	26	0.01	-0.15	0.00	-0.06	0.01	-0.01	0.18	30	0.59	-0.21	0.00	-0.06	0.00	-0.02	0.17
T	9	24	-0.54	-0.13	0.00	-0.01	0.00	-0.02	0.06	28	0.01	-0.16	0.00	-0.04	-0.01	-0.04	0.15	32	0.58	-0.20	0.00	-0.05	0.01	-0.01	0.14

G	V	FÁCIL							MEDIA							DIFÍCIL									
		PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB	PT	MED	ERR	I.MS	I.ZS	O.MS	O.ZS	PB
T	0	39	1.44	0.16	0.30	4.06	0.60	3.84	0.72	41	1.59	0.19	0.23	3.81	0.42	4.08	0.79	46	1.62	0.26	0.28	4.92	0.65	5.54	1.01
T	1	47	1.54	0.28	0.36	4.20	0.88	5.02	0.75	46	1.31	0.14	0.42	6.36	0.65	6.48	0.96	54	1.73	0.26	0.28	4.96	0.49	4.75	0.62
T	2	51	1.52	0.11	0.31	3.92	1.21	6.58	0.63	48	1.29	0.14	0.28	3.53	1.09	4.81	0.81	47	2.14	0.98	0.34	4.46	0.91	3.23	0.73
T	3	37	1.81	0.12	0.27	3.66	0.42	3.91	0.62	38	1.84	0.17	0.33	4.36	0.48	4.32	0.84	46	1.78	0.27	0.36	4.62	0.61	4.05	0.74
T	4	46	2.14	0.21	0.34	3.65	0.93	4.37	0.66	45	1.27	0.16	0.44	4.47	0.90	5.00	0.81	53	1.76	0.26	0.29	3.63	0.79	3.77	0.74
T	5	47	1.95	0.67	0.42	5.39	3.18	8.10	1.05	43	1.52	0.16	0.35	4.26	1.63	7.71	0.65	52	1.61	0.22	0.31	3.94	1.94	2.85	0.57
T	6	41	1.47	0.17	0.28	4.42	0.54	5.22	0.90	44	1.58	0.24	0.28	5.88	0.67	5.80	0.91	46	1.94	0.27	0.28	4.96	0.47	4.65	0.68
T	7	38	1.98	0.28	0.36	4.56	0.68	4.62	0.79	48	1.60	0.06	0.32	4.91	0.45	5.75	0.71	38	1.68	0.21	0.34	5.04	0.52	4.63	0.70
T	8	40	1.94	0.25	0.32	3.98	0.61	3.48	0.58	44	1.37	0.17	0.23	3.14	0.42	3.72	0.68	49	1.93	0.26	0.24	2.85	0.65	2.55	0.90
T	9	47	1.20	0.13	0.29	4.01	0.39	3.79	0.66	39	1.85	0.36	0.26	3.97	0.55	4.35	0.60	50	1.61	0.27	0.29	3.36	0.45	3.20	0.81

ANEXO G

Promedio de las diferencias entre valores del instrumento total y cada una de las pruebas (Por pruebas y habilidad del grupo)

G	V	PT	FÁCIL							MEDIA							DIFÍCIL								
			MED	ERR	LMS	LZS	O.MS	O.ZS	PB	PT	MED	ERR	LMS	LZS	O.MS	O.ZS	PB	PT	MED	ERR	LMS	LZS	O.MS	O.ZS	PB
A	0	25	1.44	0.12	0.24	4.06	0.36	3.19	0.54	27	1.11	0.05	0.17	3.50	0.22	3.40	0.53	21	0.97	0.11	0.28	4.92	0.42	5.54	0.75
M	0	12	1.00	0.02	0.24	2.80	0.27	2.86	0.56	13	0.82	0.05	0.22	2.92	0.25	3.04	0.72	12	0.98	0.10	0.19	2.06	0.26	2.00	0.60
B	0	9	0.88	0.06	0.21	2.36	0.42	2.79	0.50	11	1.59	0.19	0.18	1.66	0.39	2.69	0.61	16	1.62	0.24	0.21	1.50	0.61	2.22	0.77
A	1	31	1.03	0.24	0.27	2.56	0.54	2.88	0.56	24	0.88	0.09	0.21	4.21	0.27	4.21	0.57	29	0.95	0.08	0.18	3.43	0.29	3.17	0.40
M	1	14	1.45	0.07	0.25	4.20	0.37	4.16	0.47	11	0.66	0.05	0.42	6.25	0.65	6.31	0.96	19	1.73	0.13	0.28	4.96	0.42	4.75	0.60
B	1	14	1.16	0.06	0.30	2.99	0.71	4.52	0.53	18	1.31	0.14	0.29	2.60	0.56	3.37	0.75	19	1.52	0.22	0.21	2.11	0.43	2.48	0.44
A	2	31	1.14	0.03	0.30	3.71	0.44	3.29	0.57	27	1.02	0.02	0.15	3.17	0.21	3.52	0.64	21	1.22	0.10	0.25	4.46	0.27	3.16	0.54
M	2	14	1.19	0.03	0.29	3.80	0.71	6.28	0.58	15	1.24	0.09	0.26	3.32	0.42	4.00	0.58	13	1.03	0.13	0.24	3.54	0.63	3.13	0.45
B	2	18	1.27	0.11	0.21	1.89	1.06	4.02	0.43	19	1.20	0.13	0.27	2.53	1.09	4.80	0.67	21	2.14	0.96	0.34	2.60	0.91	2.58	0.63
A	3	26	1.14	0.11	0.20	2.89	0.29	2.56	0.48	22	0.69	0.04	0.21	3.76	0.29	3.80	0.58	24	1.12	0.16	0.20	3.68	0.30	3.93	0.53
M	3	18	1.47	0.03	0.23	3.00	0.36	3.66	0.47	13	1.58	0.15	0.28	3.33	0.38	3.73	0.65	7	1.57	0.25	0.36	4.21	0.51	3.55	0.74
B	3	15	1.62	0.08	0.25	3.03	0.39	3.48	0.56	15	1.49	0.17	0.26	2.45	0.38	2.84	0.66	17	1.41	0.23	0.23	2.12	0.61	2.21	0.58
A	4	30	1.15	0.19	0.24	3.08	0.53	2.70	0.48	28	0.85	0.07	0.23	4.22	0.33	4.62	0.53	29	1.19	0.15	0.24	3.63	0.47	3.77	0.74
M	4	16	1.31	0.02	0.32	3.31	0.48	3.78	0.59	15	1.21	0.06	0.32	3.79	0.49	4.27	0.73	11	1.28	0.18	0.26	3.22	0.58	2.84	0.59
B	4	14	1.98	0.19	0.28	2.37	0.68	3.56	0.50	16	1.27	0.16	0.39	2.54	0.80	3.77	0.66	17	1.57	0.23	0.26	2.12	0.63	2.70	0.63
A	5	25	1.95	0.64	0.25	2.48	2.02	2.25	0.78	21	1.52	0.16	0.23	2.66	0.48	2.92	0.54	23	0.97	0.06	0.25	2.58	0.51	2.56	0.48
M	5	14	1.30	0.12	0.26	3.95	1.71	8.04	0.41	14	1.20	0.03	0.23	3.50	1.48	7.24	0.45	17	1.06	0.09	0.28	3.63	1.91	2.85	0.43
B	5	15	0.72	0.04	0.38	4.71	1.17	6.17	0.75	17	0.97	0.09	0.31	3.27	1.10	4.82	0.64	18	1.30	0.16	0.27	2.48	1.57	1.77	0.44
A	6	25	1.46	0.14	0.21	4.42	0.33	4.25	0.66	30	1.16	0.04	0.27	5.88	0.30	5.80	0.72	25	1.28	0.11	0.26	4.96	0.29	4.62	0.68
M	6	15	0.91	0.03	0.24	4.18	0.36	4.54	0.63	13	1.11	0.06	0.25	3.05	0.40	3.28	0.68	15	1.60	0.24	0.23	2.75	0.30	2.52	0.57
B	6	15	0.97	0.05	0.21	2.39	0.43	3.73	0.68	13	1.58	0.24	0.25	2.12	0.67	3.07	0.77	16	1.58	0.24	0.23	2.01	0.47	2.45	0.63
A	7	22	1.75	0.25	0.26	3.33	0.49	3.05	0.53	26	1.02	0.05	0.28	4.62	0.31	4.58	0.71	17	1.28	0.08	0.27	4.01	0.34	3.91	0.54
M	7	14	1.14	0.04	0.28	3.28	0.38	3.30	0.52	17	1.31	0.05	0.31	4.63	0.43	5.29	0.71	14	1.57	0.11	0.34	5.04	0.52	4.58	0.70
B	7	13	1.20	0.04	0.34	4.56	0.52	4.62	0.77	17	1.28	0.05	0.22	2.66	0.37	3.73	0.59	18	1.35	0.18	0.22	2.27	0.45	2.50	0.58
A	8	25	1.12	0.10	0.24	3.98	0.25	2.42	0.43	31	1.06	0.04	0.15	2.61	0.18	2.00	0.42	26	0.93	0.16	0.11	2.74	0.21	1.90	0.56
M	8	11	0.87	0.03	0.26	3.11	0.61	3.48	0.51	13	0.90	0.07	0.19	1.88	0.42	3.20	0.85	11	1.58	0.25	0.17	1.85	0.30	2.31	0.55
B	8	12	1.55	0.25	0.31	2.17	0.36	1.89	0.54	11	1.21	0.17	0.18	1.45	0.35	2.24	0.57	17	1.56	0.23	0.22	1.50	0.65	2.02	0.90
A	9	30	1.16	0.05	0.28	4.01	0.36	3.79	0.56	21	0.92	0.04	0.24	3.97	0.31	4.35	0.53	30	1.10	0.11	0.28	3.36	0.31	3.03	0.81
M	9	16	1.13	0.03	0.26	2.71	0.33	3.23	0.60	13	1.16	0.08	0.24	2.72	0.38	2.94	0.58	12	1.49	0.25	0.26	3.01	0.37	2.89	0.56
B	9	16	0.93	0.06	0.20	1.70	0.37	2.56	0.48	13	1.85	0.36	0.23	2.30	0.55	2.96	0.55	16	1.52	0.24	0.23	2.24	0.45	2.27	0.67

## **Bibliografía**

- Adkins Wood, Dorothy. *Elaboración de tests*, Trillas, México, 1983.
- College Etrance Examination Borrada. *El análisis de ítems. Materiales sobre los fundamentos y prácticas de las admisiones universitarias*, The College Borrada, Hato Rey, Puerto Rico, 1979.
- Gronlund, Norman E. *Elaboración de tests de aprovechamiento*, Trillas, México, 1973.
- Muñiz F., José. *Teoría de respuesta a los ítems*, Ediciones Pirámide, Madrid, 1990.
- Rosado Chauvet, Miguel Ángel. “Prueba de aptitudes múltiples para obreros de máquinas-herramienta y ayudantes de las mismas”. Tesis para obtener el grado de Licenciado en Psicología, Facultad de Psicología, UNAM, México, 1973.
- Rosado Chauvet, Miguel Ángel. “Cinco estudios de estadística para la evaluación”. *Cuadernos del CENEVAL*, México, junio 1996-2.
- Rosado Chauvet, Miguel Ángel. “Hacia un modelo de evaluación de la docencia en la educación superior”. Tesis para obtener el grado de Maestría en Psicología Social, Facultad de Psicología, UNAM, México, 1998.
- Rosado Chauvet, Miguel Ángel. “Análisis de ítems. Teoría clásica y teoría de respuesta al ítem. Comparación y aplicación de tres modelos”. Tesis para obtener el grado de Doctor en Educación, Departamento de Ciencias de la Educación, División de Estudios de Postgrado, Universidad Autónoma de Tlaxcala, 1999.
- Rosado Chauvet, Miguel Ángel. *Metodología de investigación y evaluación*, Trillas, México, 2003.
- Tavella, Nicolás M. *Análisis de los ítems en la construcción de instrumentos psicométricos*, Trillas, México, 1978.
- Wright, Benjamin D. y Mark H. Stone. *Diseño de mejores pruebas. Utilizando la técnica de Rasch*, CENEVAL, México, 1998.